



HBR
CONSULTING

The Evolving File Analysis Marketplace

ARMA Charlotte-Piedmont and Triangle Chapters

Steve Bradley, Director
January 12, 2017



Steve Bradley
HBR Consulting
Director

sbradley@hbrconsulting.com
312-638-5123

Steve is a Director with HBR in the Information Governance group. With over 25 years of experience, Steve advises clients in the areas of information governance and compliance, records and information management, document management and litigation readiness. Steve works with legal, compliance, IT and information governance professionals to lower costs, reduce risks, ensure compliance and strategically leverage the value of their data and information resources.

Prior to HBR, Steve was a Director with Consilio and the Huron Consulting Group. He previously co-founded and served as a principal at Camden Place, and has extensive experience working within the charitable and non-profit sector, having served as a board member and/or strategic advisor to a number of organizations, addressing areas such as governance, vision crafting, mission alignment, communications, and resource development.

The Evolving File Analysis Marketplace - Agenda

- What
 - Defining File Analysis (FA)
 - The File Analysis Marketplace
 - Primary FA Use Cases
- Who
 - General Audience / Buyers / Roles
 - Representative Vendors
- Why
 - Key RIM/IG Drivers and Use Cases
- How
 - Implementation Examples
- When
 - Key Takeaways and Recommendations

What - Defining File Analysis (FA)

- File Analysis at its core involves simply the analysis of electronic file metadata and content found in unstructured data repositories
- Two Primary Levels of Analysis
 - File System Metadata
 - Includes information *about* individual files
 - Examples include contextual metadata about associated servers, volumes, shares, folders, and identity related information such as company / department / group / user permissions and ownership; as well as file specific metadata such as file owner, last author, author, file extension / item type, and create, last modified, and last accessed dates
 - File Content
 - Includes information *within* individual files
 - Represents a much more granular level of detail, and subsequently a larger data footprint and supporting set of infrastructure requirements

What - Defining File Analysis (FA)

- FA has evolved from a more distinct set of tools allowing for basic scanning, data collection, and analysis into more robust technology and platform offerings
- Actionable intelligence gleaned from detailed metadata and contextual analysis allow users to better manage and govern unstructured repositories, including:
 - Email
 - File Shares
 - ERM / EDM / ECM Systems
 - SharePoint
 - File sync and share sites such as Box.net or Dropbox
 - Data Archives
 - Business Intelligence (BI) / Data Warehouse Environments

What - The File Analysis Marketplace

- The FA Marketplace is still emerging and evolving, including more specialized niche oriented vendors, as well as established vendors offering a variety of technologies from areas such as:
 - eDiscovery
 - Information Security / Privacy / Data Loss Prevention (DLP)
 - RIM / IG / EDM / ECM
 - Data Archiving / Storage / Disaster Recovery (DR)
 - Identity and Access Management (IAM)
 - Business Intelligence (BI) / Decision Analysis
 - Big Data Analytics
- Key Source: Gartner's Market Guide for File Analysis
 - First Published in September 2014
 - Latest (2016) version: <https://www.gartner.com/doc/3446718/market-guide-file-analysis-software>

What – Primary FA Use Cases

Use Case	Description
Risk Mitigation	This includes security and risk professionals seeking to develop policies and controls for (among other things) access to data via classification according to risk or security needs.
Governance/Policy Management	FA can assist with data being classified and flowing into the appropriate platforms for governance, record keeping, compliance and analytics. In some instances, file analysis is also used as an initial step for e-discovery purposes.
Efficiency and Optimization	Organizational efficiency is based upon providing users with the data where and how they want to use it, such as in EFSS or SaaS collaboration solutions, across users and devices, and accessible from anywhere. The most common storage management use cases are to reduce the storage footprint and/or to improve storage efficiency.
Analytics	Involves presenting data in a way that provides insight to the organization for business value and anomaly detection.

Who – General Audience / Buyers / Roles

Audience / Buyer / Role	Use Case(s)
Business Intelligence	<ul style="list-style-type: none"> Analytics
Chief Analytics Officers (CAOs)	<ul style="list-style-type: none"> Analytics
Chief Data Officers (CDOs)	<ul style="list-style-type: none"> Governance/Policy Management Analytics
Chief Information Security Officers (CISOs)	<ul style="list-style-type: none"> Risk Mitigation
Data Architects	<ul style="list-style-type: none"> Governance/Policy Management Efficiency and Optimization
Data Scientists	<ul style="list-style-type: none"> Analytics
Information Stewards	<ul style="list-style-type: none"> Governance/Policy Management
IT: security	<ul style="list-style-type: none"> Risk Mitigation
IT: storage, infrastructure and operations leaders	<ul style="list-style-type: none"> Efficiency and Optimization
Legal: general counsel, chief legal officer	<ul style="list-style-type: none"> Governance/Policy Management
Litigation: support and general forensics	<ul style="list-style-type: none"> Governance/Policy Management
Risk managers, compliance and privacy authors, auditors	<ul style="list-style-type: none"> Risk Mitigation

Who – Representative Vendors Dropped From the 2016 List

Vendor Name	Product	Comments
Acaveo	Smart Information Server	Out of business as of late 2016
AvePoint	AvePoint Compliance Guardian	Identified more with InfoSec / DLP markets than FA
CommVault	Simpana	Identified more with Data Archiving / Backup markets than FA
Covertix	SmartCipher	Identified more with InfoSec market than FA
Imperva	SecureSphere File Activity Monitor	Identified more with InfoSec / Web App Security markets than FA
Microsoft (Equivio)	Zoom for Information Governance	Functionality absorbed into Office 365 eDiscovery offerings
Nuix	Nuix Information Governance, Nuix Web Review and Analytics, Nuix Sensitive Data Finder, Nuix Incident Response	Identified more with eDiscovery market than FA
Proofpoint	Proofpoint Enterprise Governance	Identified more with InfoSec / Digital Risk Monitoring markets than FA
Rational Enterprise	Rational Governance	Identified more with eDiscovery market than FA
ZyLAB	Intelligent Information Governance	Identified more with eDiscovery market than FA

Who – Representative Vendors New to the 2016 List

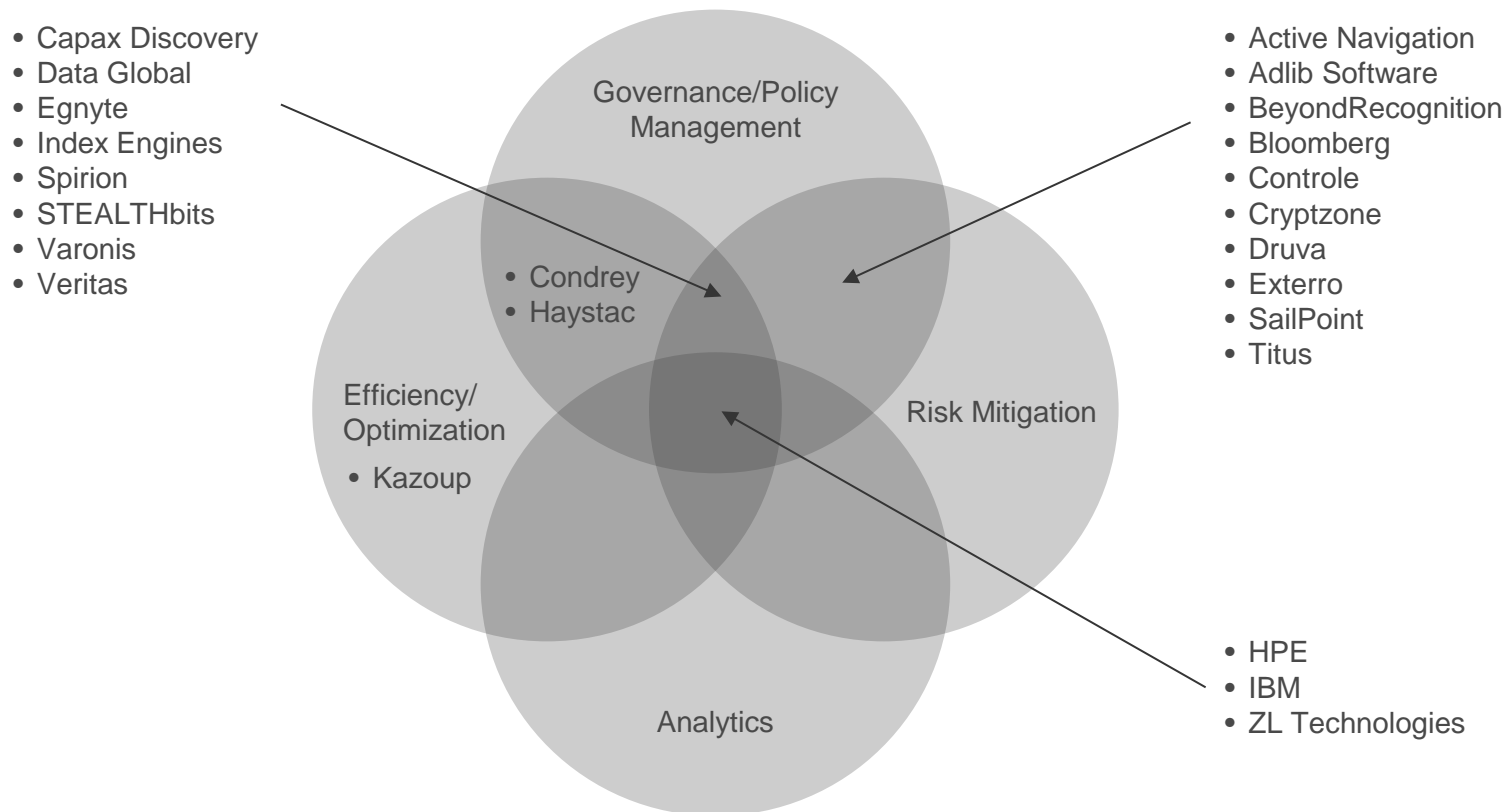
Vendor Name	Product
Capax Discovery	Enterprise Archive Solution (EAS) Data Light
Controle	eGovern
Druva	inSync
Egnyte	Egnyte Connect, Egnyte Protect
Haystac	Indago
Kazoup	Kazoup analytis platform
Titus	Titus Classification Suite, Titus Illuminate
ZL Technologies	ZL File Analysis and Management

Who – Representative Vendors On Both the 2015 and 2016 List

Vendor Name	Product
Active Navigation	Discovery Centre, Action Pack, Analysis Pack
Adlib Software	Adlib Elevate platform, PDF Enterprise
BeyondRecognition	BeyondRecognition
Bloomberg	Bloomberg Vault File Analytics
Condrey Corporation	Senergy, Galileo, Aurora
Cryptzone	Security Sheriff, Compliance Sheriff, Site Sheriff)
dataglobal	dg analyze, dg file, dg classification
Exterro	Exterro File Analytis
HPE	HPE ControlPoint, Storage Optimizer
IBM	StoredIQ
Index Engines	Catalyst
SailPoint (fka Whitebox Security)	SailPoint SecurityIQ
Spirion (fka Identity Finder)	Spirion
STEALTHbits	StealthAUDIT, STEALTHbits File System Clean-up, STEALTHbits File Activity Monitor
Veritas (formerly Symantec)	Data Insight, Information Map
Varonis	Metadata Framework

Who – Representative Vendors

Primary Use Cases Supported by 2016 List Vendors



Source: Gartner: Market Guide for File Analysis Software (19 September 2016)

Gartner's Note: Though most vendors support some elements of each use case, vendors are listed in the above diagram according to the major use case supported and what customers acquire the solution for.

Why – Key RIM/IG Drivers and Use Cases

- Key RIM/IG drivers include:
 - Risk and cost reduction
 - Increased efficiency and usability of business critical data
 - Information governance and compliance
- Key RIM/IG use cases include:
 - Assessment, identification, and discovery of unstructured content
 - Defensible disposition and ROT Removal
 - Sensitive data identification / protection
 - Data inventory, cleanup, remediation, and migration
 - Improved governance, compliance and application of controls
 - Large data set review during M&A/Divestitures and eDiscovery early case assessment (ECA)
 - Analytics and leveraging of actionable intelligence to drive decision making and define, monitor, and demonstrate success

How – Implementation Examples

Initial Discovery and Quick Win ROT Removal

- Original Scan 9.84 Terabytes / 51,256,187 Items
- **Secondary Scan 9.76 Terabytes / 45,511,422 Items**
- Two primary servers sources
 - 4.29 TB / 40,955,384 Items Total
 - 5.46 TB / 4,556,038 Items Total

Fast Facts	
File Type	Total Files
Excel Spreadsheets	1,091,654
Access Databases	6,694
PDF Documents	1,064,849
MS Word Documents	650,657
Audio Files	674,384
Video Files	30,024

- “Quick Win”
 - Server Admin team deleted directory of outdated check image files



- 5.7 million+ images, approximately 80 GB in size

How – Implementation Examples

M&A/Divestiture Large Data Set Review

•What is our primary target / goal?

- To separate / purge Company A data from Company B NetApp filers
 - Prioritizing higher risk sensitive content (e.g., PII from customers, employees, IP)

•How do we accomplish this goal?

- Demonstrating to stakeholders a good faith effort to purge all known Company A data
- Iteratively validating separation criteria logic and leveraging tool(s) to find and tag data

•How do we know when we're done?

- Ensuring 100% of filer data has been reviewed / classified and all known data purged

•Where are we now? How do we measure success?

Classification Percentages To Date

Metric	Known Targets	% Classified to Date
Server-Volume Entries	976	42.0%
Max File Entries	10,014,655,998	45.2%
Data Size	948 TB	48.5%

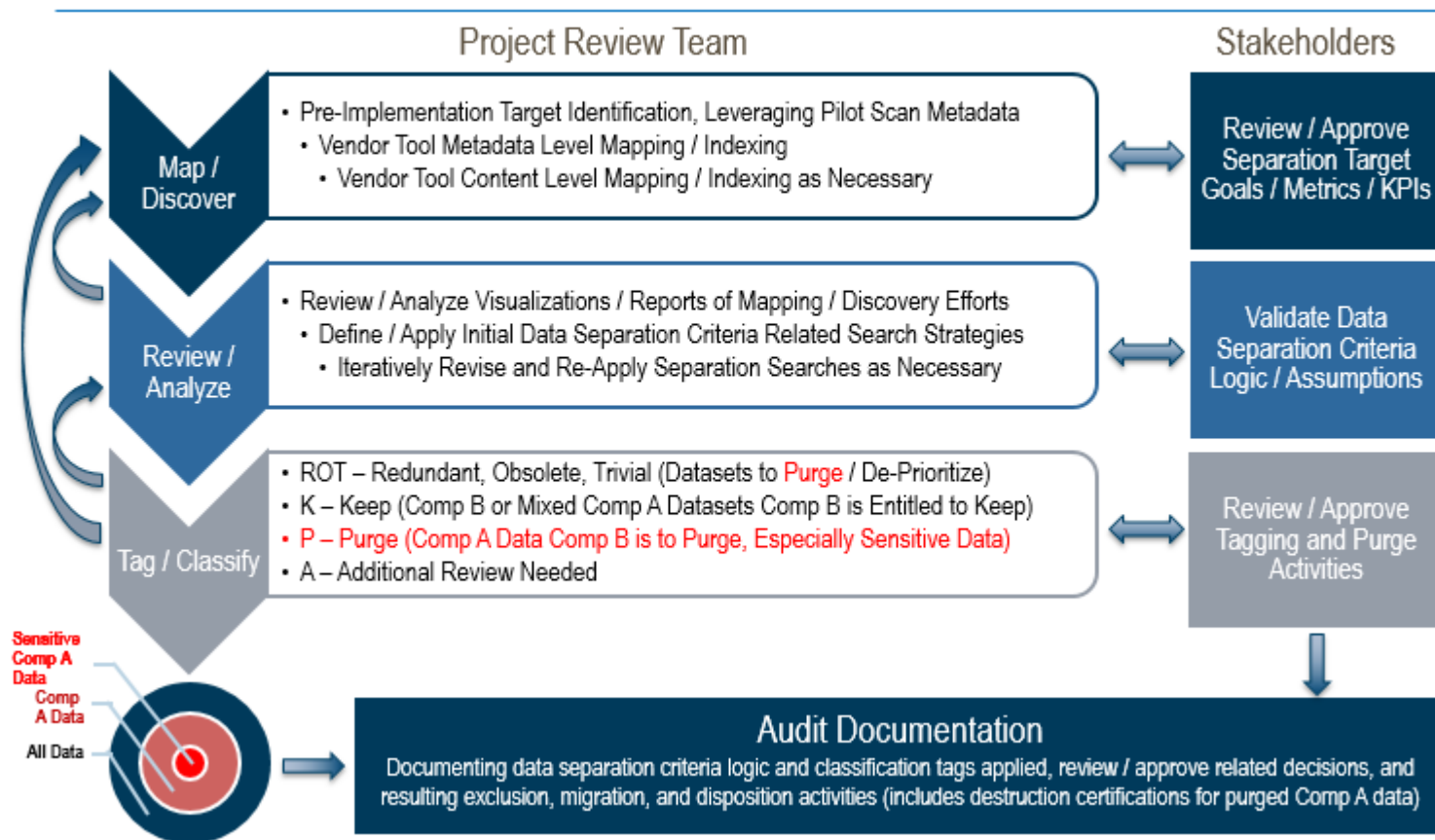
Classification Breakdown

Classification	Server-Volume	Max Files	Size InTB
TBD	566	5,485,678,988	487.80
R – Redundant	353	3,940,546,895	372.98
K – Keep	47	586,998,108	86.98
T – Trivial	9	1,432,007	0.24
O - Obsolete	1		0.00
Total	976	10,014,655,998	948.00

How – Implementation Examples

M&A/Divestiture Large Data Set Review

Group Filer Review / Data Separation Process Model



When – Key Takeaways and Recommendations

- Currently the FA Market is still emerging and evolving
- Before looking to procure a vendor solution:
 - Do a survey of your information and technology landscape
 - How much do we know about our unstructured data?
 - Key metrics include: volume of data, growth rate, age, ownership, content types
 - What tools, software, or vendor relationships do we have that can be leveraged to uncover more actionable intelligence?
 - How would our organization benefit from a FA solution?
 - Who would benefit and get behind such an effort?
 - Consider how unstructured data fits within your current policy framework and an overall content placement strategy
 - Do we have clarity/consensus on:
 - How a RRS applies to electronic data?
 - Where electronic content should be placed so proper controls may be applied?

When – Key Takeaways and Recommendations

Nuggets of Wisdom

- Driving on a paved road is easier than blazing a trail
- Test the water before diving in head first
 - Leverage free or low cost scanning / data collection options first
 - Take advantage of vendor trial / pilot offers to try before you buy
- Remember KISS and JBGE
 - Consider ways to optimize real world compliance by simplifying and clarifying RIM/IG related policies and controls
- The best business cases align with the strategic vision, mission, and priorities of your organization

Wrap Up



*Thank you for
your participation!*

Steve Bradley

Director

HBR Consulting

sbradley@hbrconsulting.com

312-638-5123